

Implications of the Genome Project for Medical Science

By

Francis S. Collins, M.D., Ph.D.
National Human Genome Research Institute, NIH

Victor A. McKusick, M.D.
Johns Hopkins University School of Medicine

Karin Jegalian, Ph.D.
National Human Genome Research Institute, NIH

Modified from an article for the Journal of the American Medical Association by Collins and McKusick

Virtually every human ailment, except perhaps trauma, has some genetic basis. In the past, doctors took genetics into consideration only in cases like birth defect syndromes and a limited set of illnesses – like cystic fibrosis, sickle cell anemia, and Huntington disease – that are caused by changes in single genes and are inherited according to predictable Mendelian rules.

Common diseases like diabetes, heart disease, cancer, and the major mental illnesses are not inherited in simple ways. But studies comparing disease risk among families show that heredity does influence who develops these conditions. As a result, many doctors are careful to ask patients about their family histories of such illnesses.

Now, with the genome project releasing a torrent of data about human DNA and promoting growing understanding of human genes, the role of genetics in medicine will change profoundly. Genetics will no longer be limited to guiding medical surveillance based on family histories, or classifying the numerous but relatively rare conditions that stem from changes in single genes.

It is true that for many of the most common illnesses, like heart disease, heredity is clearly only one of several factors that contribute to people's overall risk of developing that disease. The most common diseases in developed countries today generally arise from a complex interplay of causes, including diet, lifestyle, and environmental exposures, as well as heredity. Still, a deepening understanding of genetics will illuminate more than people's hereditary risks. Genomics reveals the basic components of cells and, ultimately, helps explain how the molecular components work together. Understanding the molecules of life and how they work will shed light on what goes wrong when diseases develop. Such detailed, fundamental understanding about our bodies will have profound effects on the ways diseases are diagnosed, on the prevention of disease, and on treatments.

Genetics in the Twentieth Century

The twentieth century saw enormous, even revolutionary, development in the field of genetics. In the spring of 1900, three different scientists brought Mendel's laws of inheritance to a wide audience. This marked the founding of genetics as a scientific discipline. In the middle of the

century, Watson and Crick revealed the chemical basis of heredity with their discovery of the double helical structure of DNA. Over the next fifteen years, scientists began to understand the role of RNA as a messenger molecule copied from DNA, and they elucidated the genetic code that allows RNA to be translated to protein.

Recombinant DNA technology burst onto the scene in the 1970s, allowing the preparation of pure samples of short DNA segments. Sequencing DNA was still brutally difficult, until Sanger and Gilbert independently developed methods to sequence DNA in the late 1970s. Remarkably, all genome projects have used the same basic technology for DNA sequencing that Sanger developed, though there have been major advances in automating the analysis of sequencing results in the last 15 years.

In 1980 scientists began mapping genes whose variants cause disease. In 1983, for example, mapping localized the Huntington disease gene to chromosome 4. But even after mapping them, finding the genes actually responsible for diseases remained an arduous task. Years of work were required to develop detailed maps over the regions containing long-sought genes, and then to search among the genes in these areas to find the ones specifically desired.

Many researchers longed for a more systematic way of approaching the genome. Meanwhile, some scientific leaders, particularly in the Department of Energy, began touting the possibility of organizing an effort to sequence the entire human genome. In the late 1980s much controversy raged about such proposals; many scientists expressed concern that such a project was technologically impractical and, if launched, likely to consume vast amounts of funding that could go to other kinds of scientific research instead. But with the strong support of a panel of the National Academy of Sciences and the enthusiasm of a few leaders in the U.S. Congress, in 1990 the National Institutes of Health and the Department of Energy initiated the U.S. portion of the Human Genome Project (HGP).

The Human Genome Project

The HGP struck the research community as an audacious undertaking; when the project began, DNA sequencing technology was not yet in a position to tackle the three billion base pairs of the human genome. To take on a project of such magnitude, the leaders of the project, headed by James Watson, decided to develop a detailed set of plans and to define intermediate milestones. Before attempting wholesale sequencing, they decided to develop maps of the genome. The maps would not only help build a scaffold for the eventual sequencing; they would themselves be of great value to scientists hunting for disease genes.

The HGP plan included the decision to map and sequence the genomes of other organisms that have been important to the study of biology: bacteria, yeast, roundworm, fruit fly, and mouse. In addition, the project sought to improve sequencing technology. Perhaps most unusual of all for a scientific project, from the outset, three to five percent of the HGP budget funded research on the ethical, legal, and social implications (ELSI) of having so much new genetic information about our species. In the past, analysis of the ethical, legal, and social consequences of a scientific revolution often lay dormant until a crisis developed. This time, those who started the HGP hoped to inspire ethicists, social scientists, legal scholars, theologians, and others to turn their attention early to the dilemmas likely to arise when knowledge about the genome increased – from the possibility of genetic discrimination to more philosophical issues like the relative importance of genetics in determining who we are.

From its inception, the HGP has been an international effort. The United States has made the largest investment, but important contributions have come from many countries, including Britain, France, Germany, Japan, China, and Canada. When the project began, the complete

human genome sequence was expected by the year 2005, though there was certainly very little reason to be confident then that this goal could be achieved. But one by one, the intermediate milestones were accomplished.

The HGP participants had agreed all along to release all maps and all DNA sequence data into public databases. With access to increasingly detailed maps of the genome, the research community began to identify genes involved in diseases more and more quickly. While less than 10 genes had been identified by the technique known as positional cloning in 1990, that number grew to more than 100 by 1997.

By 1996, with complete genome sequences obtained for several species of bacteria and for yeast, HGP participants decided to attempt sequencing human DNA, at least on a trial scale. The availability of new kinds of sequencing machines and the effort by a newly formed private company to sequence the human genome further spurred the effort. By 1999, confidence grew that HGP participants were ready to sequence the three billion base pairs of the human genome. In June 2000, both the private company and the Human Genome Project's international consortium announced the completion of "working drafts" of the human genome sequence.

Current Genomic Research

The working draft sequence covers 90% of the human genome. Though it represents a major milestone, a vast amount of work remains to understand the genome.

- The human genome must be sequenced completely. Gaps and ambiguities that remain in the draft sequence must be clarified. This finishing process had been accomplished for chromosomes 21 and 22 by the summer of 2000, and will be carried out for the rest of the genome by 2002.

- Genome sequences will be obtained for other organisms. Comparing genome sequences from different species will be a great aid in revealing the genes, since the stretches of DNA that code for protein and the regions in genes that regulate their expression tend to be conserved among species. Large-scale sequencing of laboratory mouse DNA has already started. Projects to sequence rat and zebrafish DNA will not be far behind. Scientists in both the public and private sectors are seriously considering sequencing other large vertebrates' genomes, including those of the pig, dog, cow, and chimpanzee.

- An intense effort is in progress to develop a catalog of human DNA variations. While our DNA sequences are 99.9% identical to each other, the 0.1% of variation is expected to contain many clues about the genetic risk for illnesses. A partnership between several private companies and the government has begun to compile a catalog of common variations; the consortium will find approximately 1 million variations (specifically, single nucleotide polymorphisms or SNPs) by early 2001.

- Technology is being developed to study the expression of many genes at once. The new methods are beginning to allow researchers to observe in single experiments whether as many as 10,000 genes are turned on or off in various cells or under different conditions. Such studies help reveal how different tissue types differ and what alterations in gene expression accompany the development of diseases. Such analyses have already helped find differences among tumors that otherwise seemed identical.

- Researchers have had practice in large-scale analysis of DNA and RNA. Now, they are beginning to study proteins on a sweeping scale. These studies will look at the amounts of

different proteins, where they're located in the body and in cells, how they may be modified chemically after translation, and how they interact with other molecules.

- Because large-scale experiments are flooding researchers with information, a field called computational biology is emerging and will become increasingly important for analyzing all this information.
- The ELSI research program has so far concentrated particularly on issues of privacy, genetic discrimination, and education. It is starting to give attention to the implications to medicine and to society of having increased information about human variation.

Medical Research in the Twenty-First Century

Obtaining the sequence of the human genome represents the end of a first stage in a long process toward understanding the makeup of life. For medicine to take full advantage of the advances in genetics, major challenges lie ahead.

- Having the human genome sequence and knowing the DNA spelling variations among people will help reveal which genes contribute to the risks for common diseases. This will be a challenging task. For diabetes, for example, researchers expect that five to ten – and perhaps more – genes are involved, all of which have forms that increase the risk for disease slightly. Those genes interact with each other and the environment in complex ways. Finding a gene involved in such diseases is many times harder than in cases where a disease stems from variations in a single gene.

Even so, researchers are optimistic that by precisely diagnosing different forms of diseases like diabetes, heart disease, and cancer and by developing a large catalog of genetic variations, they will begin to find genes for some of the most common illnesses in the next five to seven years.

- Researchers will increasingly bring to light the molecular processes that normally maintain the human body in good working order. The studies will also reveal how the same processes are disturbed in illness. Now, our understanding of the molecular basis of most common diseases is quite limited. The discovery of each gene that affects the risk for an illness will reveal a clue about how that illness arises.
- Researchers will need to develop and apply methods that analyze many drugs at a time for their potential affect on disease-related genes and gene products. The pharmaceutical industry has been gearing up for this opportunity, and most pharmaceutical companies now expect that the majority of future drug development will come from the field of genomics. New, efficient ways of analyzing the effects of many drugs should identify those that block or stimulate particular genetic pathways. A gratifying recent example is the development of the drug STI-571, which was designed to block the activity of the *bcr-abl* gene. The protein product is produced when a rare fusion occurs between chromosomes 9 and 22. This fusion is characteristic of, and the central cause of, chronic myelogenous leukemia (CML). STI-571 blocks the ability of the *bcr-abl* product to chemically modify an unknown molecule; the drug has shown dramatic results in early clinical trials on patients with advanced CML.
- Genomics is likely to help allow the prediction of individuals' responsiveness to particular drugs, since variations in drug response often stem from genetic differences. For example individuals break down particular drugs at different rates. Researchers are beginning to correlate variations in the spellings of genes with variations in responsiveness to different drugs. This new field is called pharmacogenomics and promises to make prescription of drugs a much more individualized affair in the future.

- The field of gene therapy has sustained a series of disappointments over the past few years and has gone back to wrestling with basic scientific questions. Safer and more effective ways of transferring genes into the human body must be developed before gene therapy can live up to its promise and play a significant role in the treatment of disease.

Genetics in the Medical Mainstream

Over the next quarter century, the practice of medicine will increasingly depend on an understanding of molecules and genetics.

By the year 2010, predictive genetic tests are likely to be available for many common conditions, allowing individuals who wish to know this information to learn what their individual susceptibilities are, and to take steps to reduce those risks for which interventions are available. The interventions could take the form of medical surveillance, life style modifications, changes in diet, or drug therapy. For example, those at highest risk for colon cancer could undergo frequent colonoscopies for screening, which would prevent many premature deaths. Predictive genetic tests are likely to be applied first in cases where individuals have a strong family history of a particular condition; in fact, such testing is already available for a few conditions, including breast cancer and colon cancer.

But with increasing genetic information available about common illnesses, this kind of genetic risk assessment will become more generally available. Many primary care providers will need to practice genomic medicine; they will need to explain complex statistical risk information to healthy patients who are seeking to maximize their chances of staying well. This will require substantial advances in the understanding of genetics by health care providers. Another crucial step is the passage of legislation that bans the use of genetic information that predicts future risk in decisions about health insurance and employment. Individuals should not have to forgo acquiring genetic information about themselves out of fear of discrimination. Although more than two dozen states have taken some action on the issues of genetic privacy and genetic discrimination, an effective Federal law would help eliminate the patchwork of different levels of protection across the U.S.

By 2020 the impact of genetics on medicine will be even more widespread. The pharmacogenomics approach for predicting drug responsiveness will be standard practice for many drugs. New gene-based "designer drugs" will be coming on the market for diabetes, hypertension, mental illness, and a long list of other conditions. The diagnosis and treatment of cancer will likely be transformed. By 2020, it is likely that every tumor will undergo precise molecular fingerprinting, to catalog the genes that have gone awry, and therapy will be individually targeted to that fingerprint.

Despite these hopeful and expectant projections, certain tensions will also exist. Access to health care is already a major problem, and our medical care system has to change in significant ways to spread the benefits of the new advances. Anti-technology movements, already active in the U.S. and elsewhere, are likely to gather momentum as the focus of genetics turns even more intensely on ourselves. Though the benefits of genetic medicine will be profound, there will be those who consider this unnatural and dangerous. Efforts at public education need to start now to explain the potential benefits and to be honest about the risks.

Summary

In conclusion, we face a time of dramatic change in medicine. As we cross the threshold of the new millennium, we simultaneously cross a threshold into an era where the human genome sequence is largely known. We must commit ourselves to exploring the application of these

powerful tools to the alleviation of human suffering, a mandate that undergirds all of medicine. At the same time, we must be mindful of the great potential for misunderstanding of a field that is developing very quickly, and make sure to advance the ethical and legal consideration of genetics with just as much vigor as the medical research.